ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Integrative Analysis of Metabolome and Transcriptome Reveals the Mechanism of Flavonoid Biosynthesis in *Lithocarpus polystachyus* Rehd

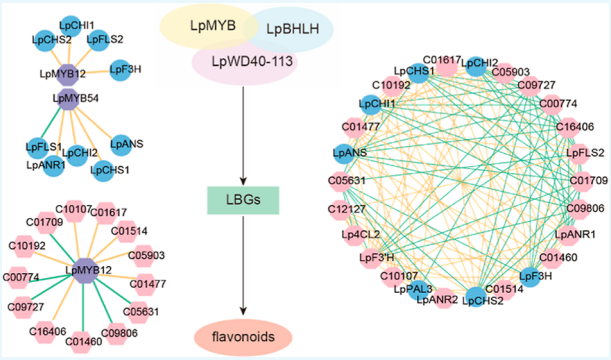
Duoduo Zhang, Shuqing Wang, Limei Lin, Jie Zhang, Minghui Cui, Shuo Wang, Xuelei Zhao, Jing Dong, Yuehong Long,* and Zhaobin Xing*

Cite This: *ACS Omega* 2022, 7, 19437−19453

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: *Lithocarpus polystachyus* Rehd has received great attention because of its pharmacological activities, such as inhibiting oxidation and lowering blood glucose and blood pressure, and flavonoids are one of its main pharmacodynamic components. It is important to understand the mechanisms of the flavonoid biosynthetic pathway of *L. polystachyus*, but the regulation of flavonoid biosynthesis is still unclear. In this study, differentially expressed genes and differentially accumulated metabolites in *L. polystachyus* were studied by integrating transcriptomics and metabolomics technologies. We confirmed the key genes involved in the flavonoid biosynthesis of *L. polystachyus*, including *LpPAL3*, *LpCHS1*, *LpCHS2*, *LpCHI2*, and *LpF3H*, which had consistent expression patterns with their upstream and downstream metabolites, and there is a significantly positive correlation between them. Compared to mature leaves, stems and young leaves are higher in the expression levels of key structural genes. We deduced that the MYB and bHLH transcription factors regulated the biosynthesis of different flavonoid metabolites and their regulatory patterns. Among them, *LpMYB2*, *LpMYB20*, *LpMYBS4*, *LpMYB12*, and *LpWD40-113* positively regulated the biosynthesis of flavones and flavanones. This discovery preliminarily revealed the pathways and key genes of flavonoid biosynthesis in *L. polystachyus*, which provided a reference for further study on flavonoid biosynthesis.



INTRODUCTION

Lithocarpus polystachyus Rehd, also known as sweet tea, is a kind of evergreen tree of the genus Fagaceae, which is widely distributed in southeast China.[1] Dihydrochalcone, a kind of flavonoid and natural sweetener, is the main active component in *L. polystachyus*. The dihydrochalcone content of *L. polystachyus* was significantly higher than other species. The content of phlorizin in sweet tea is 100 times higher than that in apples, suggesting that sweet tea is an excellent natural source of phlorizin.[2] With pharmacological activities such as inhibiting oxidation and lowering blood glucose and blood pressure, *L. polystachyus* has been widely studied. Flavonoids are one of the main pharmacodynamic components of *L. polystachyus*.[3,4] They are widely distributed in plants, with various biological functions, such as resisting biotic and abiotic stresses, regulating phytohormone activity, and so forth.[5] Flavonoids are also one of the main active ingredients in many medicinal plants, with pharmacological activities such as treating cancer, inflammation, and cardiovascular diseases.[6] Therefore, it is significant to understand the biosynthesis mechanism of flavonoids in *L. polystachyus*.

Flavonoids can be divided into six categories according to their structures: flavanones, flavones, isoflavones, chalcones, flavonols, and anthocyanins (Figure 1).[7] Various genes and enzymes regulate flavonoid biosynthesis. Phenylalanine is transformed via the phenylpropanoid pathway into coumaroyl-CoA that then enters the flavonoid biosynthesis pathway to produce chalcone. Chalcone undergoes intramolecular cyclization to yield flavanone, the main precursor of other flavonoids. From this flavanone, the pathway diverges into different side branches to form different types of flavonoids by specific enzymes.[8]

In addition to structural genes, some transcription factors (TFs) such as R2R3-MYB, bHLH, and WD40 can control multiple enzymatic steps in the flavonoid biosynthetic pathway alone or in collaboration with other factors. Early biosynthesis genes in the flavonoid biosynthetic pathway are mainly activated by independent R2R3-MYB, which leads to the

Received: February 24, 2022
Accepted: May 23, 2022
Published: May 31, 2022

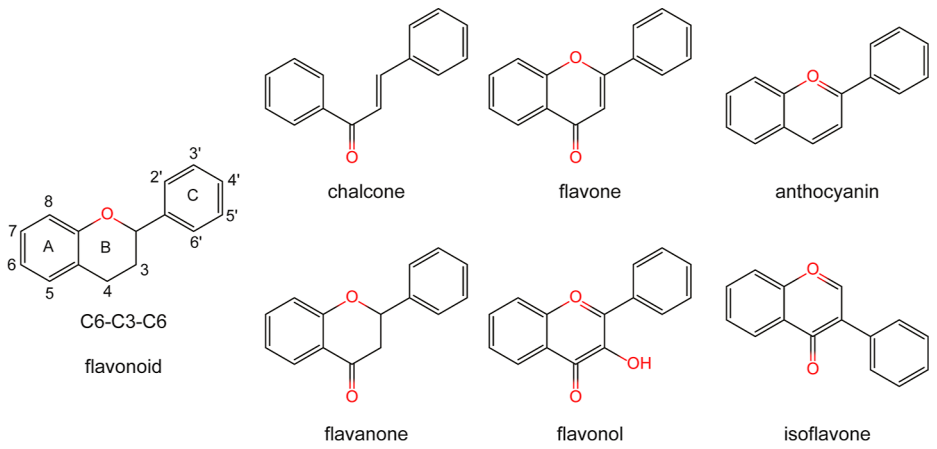


Figure 1. Chemical structures of flavonoids.

production of flavonols, while the activation of late biosynthesis genes (LBGs) usually requires a ternary complex of TFs (MYB-bHLH-WD40, MBW).[9] *PAP1 (MYB75)-TT8/GL3-TTG1 (WD40)*, *PAP2 (MYB90)-TT8/GL3-TTG1*, *MYB113 (PAP3)-TT8/GL3-TTG1*, and *PAP4 (MYB114)-TT8/GL3-TTG1* have been shown to activate the expression of LBGs, such as *AtDFR*, *AtANS*, and *AtANR* in *Arabidopsis*.[10] MdMYB1, MdMYB10, MdMYB12, and MdMYB22 regulate flavonoid biosynthesis in apple (*Malus domestica*), with the former two regulating anthocyanin biosynthesis in a tissue-specific manner and the latter two controlling flavonol and proanthocyanidin (PA) biosynthesis.[11] Some MYBs negatively regulate flavonoid biosynthesis. The overexpression of *FaMYB1* in strawberries seriously affects the expressions and enzyme activities of late flavonoid biosynthesis genes.[12] In addition, bHLH is an important class of TFs, belonging to the MYC family, with a structure of the helix−loop−helix domain. *MdbHLH3* binds to the promoter of *MdDFR* to activate its expression in apple.[13]

Previous studies have shown that the content of flavonoids is different in different organs and at different growth and development stages,[14,15] but they did not clarify the regulation mode of flavonoid biosynthesis in *L. polystachyus*. This study analyzed the differentially accumulated flavonoids (DAFs) and differentially expressed genes (DEGs) in different organs and at different growth and development stages by integrating transcriptomics and metabolomics technologies. This research aims to elucidate the key genes and the regulatory relationship between TFs and key genes and to analyze the mechanism of flavonoid biosynthesis in *L. polystachyus*.

MATERIALS AND METHODS

Experimental Materials. The leaves and stems of wild plants at different growth and development stages were selected in Bama County, Guangxi, China. According to the growth status of the plants, the mature leaf samples of a 3 year-old *L. polystachyus* were named LM, the stem samples SM, the leaf samples at the developing stage as LD, and the leaf samples at the early stage as LY. Each sample was subjected to three independent biological replicates, and the samples were frozen in liquid nitrogen and stored at −80 °C for subsequent experiments.

Widely Targeted Metabolome Analysis. A widely targeted metabolome analysis was performed by Metware Biotechnology Co., Ltd. (Wuhan, China). The freeze-dried samples were crushed using a mixer mill (MM 400, Retsch, Dusseldorf, Germany) with a zirconia bead for 1.5 min at 30 Hz. The sample powder (0.1 g) was fully dissolved in 0.6 mL of 70% aqueous methanol, which was followed by extraction using an ultrasonic power of 300 W with 5 s breaking and 8 s intermittent time for 30 min and extracted overnight at 4 °C. After centrifugation (10,000*g*) for 10 min, the extracts were filtered using a 0.22 μm filter membrane (SCAA-104, ANPEL, Shanghai, China) for ultraperformance liquid chromatography (UPLC)−tandem mass spectrometry (MS/MS) analysis.

An UPLC-ESI-MS/MS system (UPLC, Shim-pack UFLC SHIMADZU CBM30A system; MS, Applied Biosystems 4500 Q TRAP) was used to extract the samples. The gradient elution solvents comprised mobile phase A (pure water with 0.04% acetic acid) and mobile phase B (acetonitrile with 0.04% acetic acid). The gradient program conditions included 0 min, 95% A; 10 min, 5% A; 11 min, 5% A; 12 min, 95% A; 12.1 min, 95% A; 15 min, 95% A. All samples were analyzed using an ACQUITY UPLC HSS T3 C18 column (1.8 μm, 2.1 mm × 100 mm, Waters). The temperature was maintained at 40 °C. The flow rate and the injection volume were 0.35 mL/min and 4 μL, respectively.

An API 4500 Q TRAP MS system was equipped with electrospray ionization (ESI) and Turbo ion-spray interfaces operating in positive and negative ion modes and controlled using Analyst 1.6.3 software. The ESI source operation parameters included an ion source, turbo spray (550 °C); ion spray voltage at 5500 V; ion source gas I (GSI) at 50 psi; gas II (GSII) at 60 psi; curtain gas (CUR) at 30 psi. The collision gas (CAD) was set to high, and triple quadrupole (QQQ) scans were acquired via multiple reaction monitoring (MRM) experiments with CAD (nitrogen) at 5 psi. The declustering potential and collision energy for individual MRM transitions were further optimized. We monitored a specific set of MRM transitions based on the metabolites eluted within each period.

The MS data were processed using Analyst 1.6.3 software to obtain the total ion flow current and MRM detection of multimodal maps of mixed mass control samples. The horizontal coordinate is the retention time (Rt) for metabolite detection, and the vertical coordinate is the ion flow intensity (cps, count per second) for ion detection. Based on the self-built metware database (MWDB), MassBank (http://www.massbank.jp/), and METLINE (https://metlin.scripps.edu/) databases, material characterization was carried out according

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

to the information of the secondary spectrum. The MRM detection of multimodal maps shows the substances that can be detected in a sample, with each differently colored MS peak representing a metabolite detected. The signal intensity (CPS) of the characteristic ions is obtained in the detector by screening each substance with a triple quadrupole. The Analyst 1.6.3 software was used to process the MS data, integrate and correct chromatographic peaks, and export the integration data of the chromatographic peak area for preservation.

Metabolites were subjected to principal component analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA), and the variable importance in a project (VIP) of the OPLS-DA model was obtained. The significantly differentially accumulated metabolites (DAMs) were screened according to the criteria of VIP $\geq$ 1, fold change $\geq$ 2, or fold change $\leq$ 0.5.

Determination of Total Flavonoid Content. After the samples were dried to a constant weight and ground into powder, 100 mg of powder was weighed and extracted with 10 mL of 70% aqueous methanol, which was followed by extraction using an ultrasonic power of 300 W with 5 s breaking and 8 s intermittent time for 30 min. They were then centrifuged at 12,000 rpm for 10 min, and the supernatant was taken as the sample for testing. After the supernatant was diluted three times, the content of total flavonoids was determined by the spectrophotometric method according to the instructions of the kit to determine total flavonoids in plants (Solarbio, Beijing, China) and was calculated by drawing a standard curve with rutin as the standard.

Quantification of Flavonoid Metabolites. The supernatant was filtered by a microporous membrane (0.22 μm pore size) and stored in an injection vial for UPLC-MS/MS analysis. A total of 10 flavonoid metabolites were randomly selected to verify the data accuracy and reliability of the metabolomics analysis, including 2 flavanones (phloretin and naringenin), 2 flavones (apigenin and luteolin), 2 flavonols (kaempferol and myricetin), 1 flavanols (epicatechin), 2 isoflavones (genistein and biochanin A), and 1 anthocyanin (cyanidin-3-*O*-glucoside). We quantified the content of 10 flavonoid metabolites in *L. polystachyus*. The standard solution of 10 flavonoid metabolites was prepared and used to draw the standard curve. A mixture of 223 mg of luteolin, 290 mg of phloretin, and 314 mg of apigenin was dissolved in a 70% ethanol solution at a constant volume of 25 mL, which was named Mix1. A total of 540 mg of cyanidin-3-*O*-glucoside, 472 mg of myricetin, and 354 mg of genistein were mixed and dissolved in a 70% ethanol solution at a constant volume of 25 mL, and the mixture was named Mix2. A total of 400 mg of epicatechin, 345 mg of naringin, 465 mg of kaempferol, and 408 mg of biochanin A were dissolved in a 70% ethanol solution at a constant volume of 25 mL, which was named Mix3. The above mixture was diluted 5-fold, 10-fold, 20-fold, and 50-fold in gradient and then detected by the UPLC system with the original standard solution and sample supernatant. The standard curve of the obtained standard substance was used to calculate the solubility of the compounds in the samples (luteolin: $Y = 2.92 \times 10^5X - 1.90 \times 10^4$, $R^2 = 0.999604$; phloretin: $Y = 5.85 \times 10^5X - 4.16 \times 10^4$, $R^2 = 0.999627$; apigenin: $Y = 4.39 \times 10^5X - 3.60 \times 10^4$, $R^2 = 0.999848$; cyanidin-3-*O*-glucoside: $Y = 5.08 \times 10^4X - 2.23 \times 10^4$, $R^2 = 0.988916$; myricetin: $Y = 1.64 \times 10^5X - 7.10 \times 10^4$, $R^2 = 0.997226$; genistein: $Y = 4.45 \times 10^5X + 3.41 \times 10^4$, $R^2 = 0.999925$; epicatechin: $Y = 4.62 \times 10^4X - 4.92 \times 10^4$, $R^2 = 0.995755$; naringin: $Y = 5.61 \times 10^5X + 1.80 \times 10^3$, $R^2 = 0.999890$; kaempferol: $Y = 2.27 \times 10^5X - 1.60 \times 10^5$, $R^2 = 0.998497$; biochanin A: $Y = 4.49 \times 10^5X + 1.07 \times 10^4$, $R^2 = 0.999968$). The gradient elution solvents included mobile phase A (water with 0.1% formic acid) and mobile phase B (methanol). The gradient procedure is shown in Table 1. The samples were separated on a C18 column (ACQUITY UPLC BEN C18 column: 1.7 μm, 2.1 mm × 50 mm) at a flow rate of 0.25 mL/min. The injection volume was 4 μL, with a temperature of 40 °C and a detection wavelength of 280 nm.

Table 1. Sample and Standard Gradient Elution Procedure

Mix1			Mix2			Mix3		
time (min)	A (%)	B (%)	time (min)	A (%)	B (%)	time (min)	A (%)	B (%)
0	95	5	0	95	5	0	95	5
10	30	70	15	20	80	20	40	60
11	30	70	16	20	80	21	40	60
12	95	5	17	95	5	22	95	5
15	95	5	20	95	5	25	95	5

Total RNA Extraction and RNA-Seq Analysis. The total RNA was extracted and purified using an RNAprep Pure Plant Plus Kit (polysaccharide and polyphenol-rich) (TIANGEN, Beijing, China). A Nanodrop One (Thermo Fisher Scientific, Waltham, MA, USA) spectrophotometer and Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) were used to determine the purity, concentration, and quality of total RNA. The mRNA library was constructed using RNA (3 μg) from each sample and then sequenced on an Illumina NovaSeq 6000 platform. Adaptor sequences and low-quality reads were removed from raw reads. Clean data was obtained after filtering. The Q20, Q30, and GC content in the clean data were calculated. Trinity software (v2.8.5) was adopted to assemble sequences using clean data. Low-expression transcripts were filtered to construct unigenes. Transcriptome sequencing was performed by using an Illumina HiSeq high-throughput sequencing platform. RNA-seq generated 40.37−49.20 M clean readings and 6.05−7.38 Gb clean bases. The Q30 percentage (sequencing error rate less than 0.1%) was over 92%, and the GC content was all around 44%.

The unigenes were compared and annotated with NR, Swiss-Prot, Gene Ontology (GO), euKaryotic Ortholog Groups (KOG), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases using the BLAST software (v2.2.31) with default parameters. TransDecoder software (V3.0.0) was applied to predict the unigene coding sequence and amino acid sequence of the unigenes. The iTAK software was used to predict the TFs. The gene expression levels in each sample and fragments per kilobase of transcript per million mapped reads (FPKM) were estimated using RSEM. Differential expression analysis of different organ comparison groups and the comparison groups at different growth and development stages was performed using the DESeq2 (v1.10.1) software R package. Fold change $\geq$ 2 and false discovery rate (FDR) < 0.01 were defined as DEGs. GOSeq (v2.12) and KOBAS (v2.0) software were used for GO and KEGG pathway functional enrichment analyses of the DEGs. The genes, TFs, and transporters were identified through annotation information of NR, Swiss-Prot, and GO.

Combined Transcriptome and Metabolome Analysis. The difference multiples of DAMs and DEGs were counted,

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

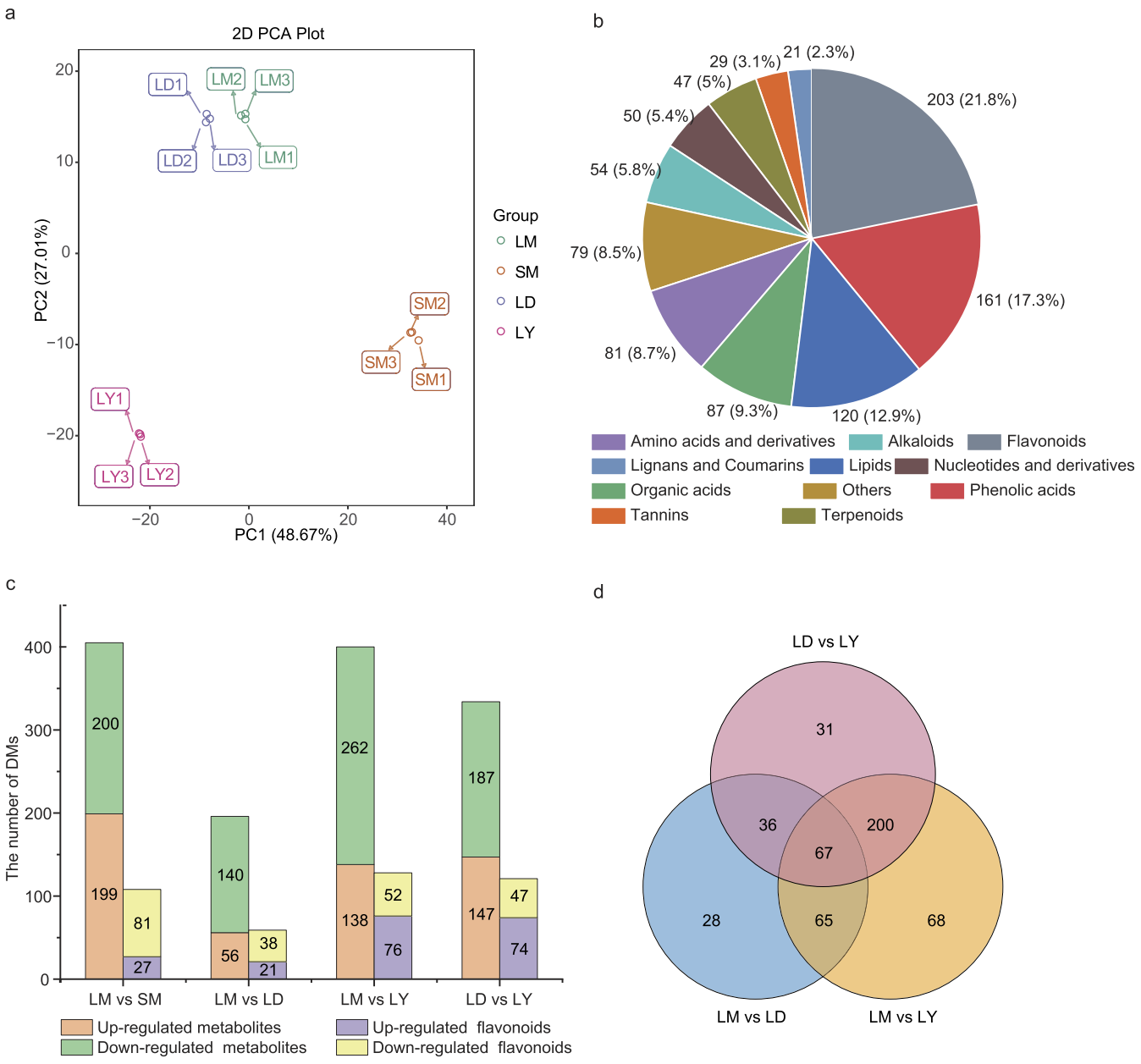


Figure 2. Analysis of metabolites in different organs and at different growth and development stages of *L. polystachyus*. (a) PCA of metabolic groups. (b) Metabolite classification. (c) Number of DAMs and DAFs in different organs and at different growth and development stages. (d) Venn diagram of DAMs at different growth and development stages.

the Pearson correlation coefficient (PCC) was calculated by using the R package, and PCC was used to measure the correlation between DAMs, between DEGs, and between DAMs and DEGs. PCC ≥ 0.8 and $p < 0.01$ were considered to have a significant correlation. Metabolome and transcriptome relationships were visualized by using Cytoscape (version 3.7).

Quantitative Real-Time Polymerase Chain Reaction. To verify the accuracy of the expression levels obtained from RNA-Seq analysis, 20 DEGs were randomly selected for quantitative real-time polymerase chain reaction (qRT-PCR). Primers specific for the selected DEGs were designed by Primer 5.0 (Table S1), and specificity was identified by dissolution curve analysis. qRT-PCR was performed on Applied Biosystems 7900HT (Thermo Fisher Scientific, Waltham, MA, USA) using Talent qPCR PreMix (SYBR Green) (TIANGEN, Beijing, China), and the *GAPDH* gene was used as an internal reference gene.[16] The amplification reaction conditions were as follows: predenaturation at 95 °C for 3 min was followed by 40 cycles at 95 °C for 5 s and 60 °C for 15 s. All qRT-PCR experiments were performed in three biological replicates, and the relative expression levels were calculated based on the $2^{-\Delta\Delta Ct}$ method.

RESULTS

Analysis of Metabolites of *L. polystachyus*. To compare the DAMs of different organs and different growth comparison groups and development stage comparison groups in *L. polystachyus*, the UPLC-ESI-MS/MS system was used to analyze the metabolome of the samples. The results of PCA showed that PC1 was 48.67% and PC2 was 27.01%, indicating that the four samples were separated and there was a large difference between groups, but there was no difference within

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

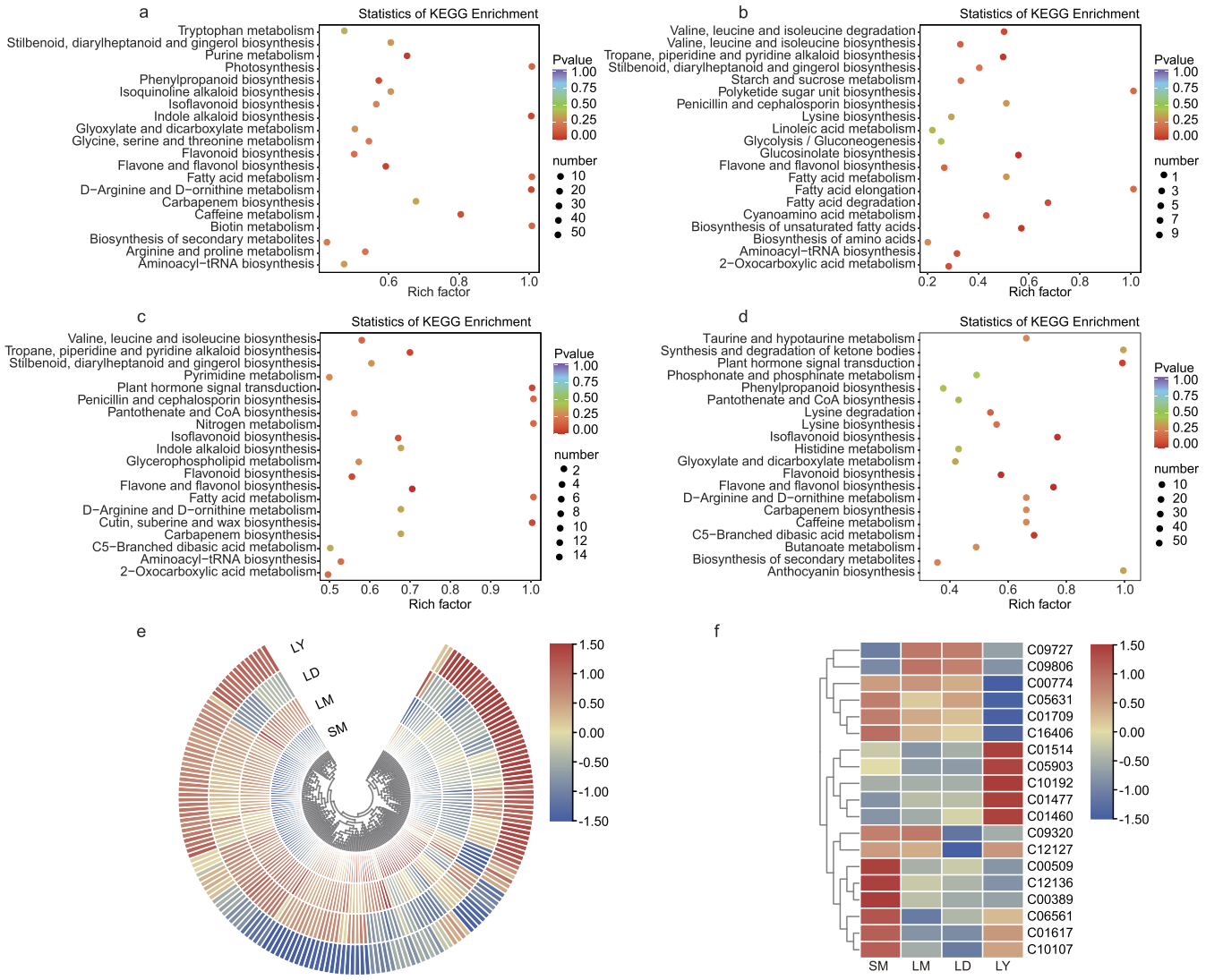


Figure 3. KEGG enrichment and cluster analysis of DAMs of *L. polystachyus*. (a) KEGG enrichment of DAMs in the "LM versus SM" group. (b) KEGG enrichment of DAMs in the "LM versus LD" group. (c) KEGG enrichment of DAMs in the "LM versus LY" group. (d) KEGG enrichment of DAMs in the "LD versus LY" group. (e) Cluster analysis of DAFs. (f) Cluster analysis of DAFs mapped to the Ko00941 pathway.

groups (Figure 2a). A total of 933 metabolites were identified in the four samples, with the largest proportion being flavonoids (203, 21.78%), followed by phenolic acids (161, 17.27%), lipids (120, 12.88%), organic acids (87, 9.33%), and amino acids and their derivatives (81, 8.69%), and lignans and coumarins (21, 2.25%) accounted for the least (Figure 2b). The flavonoids identified in the samples included 11 chalcones, 13 flavanones, 5 dihydroflavonols, 2 anthocyanins, 92 flavones (79 flavones and 13 flavone *C*-glycosides), 49 flavonols, 17 flavanols, and 14 isoflavones.

With VIP $\geq$ 1 and fold change $\geq$ 2 or fold change $\leq$ 0.5 as the criteria, a total of 658 DAMs were screened, of which 174 were flavonoids. In the "LM versus SM" group, there were 405 DAMs, of which 109 DAFs and 199 DAMs were upregulated, and 27 DAFs were highly accumulated in SM. In the "LM versus LD" group, there were 196 DAMs, of which 59 DAFs and 56 DAMs were upregulated, and 21 DAFs were highly accumulated in LD. In the "LM versus LY" group, there were 400 DAMs, of which 128 DAFs and 138 DAMs were upregulated, and 76 DAFs were highly accumulated in LY. In the "LD versus LY" group, there were 334 DAMs, of which 121 DAFs and 147 DAMs were upregulated, and 74 DAFs were highly accumulated in LY (Figure 2c). In the comparison groups at different growth and development stages, there were 67 common DAMs, of which 35 were DAFs (Figure 2d).

All DAMs were annotated to the KEGG pathway for enrichment analysis, and the results showed that they were primarily enriched in the "flavone and flavonol biosynthesis" pathway and the "flavonoid biosynthesis" pathway of all comparison groups. The DAMs were primarily enriched in the "isoflavonoid biosynthesis" pathway of the "LM versus SM," "LM versus LY," and "LD versus LY" groups. In addition, they were also generally enriched in the "phenylpropanoid biosynthesis" pathway of the "LM versus SM" and "LD versus LY" groups (Figure 3a−d).

Cluster analysis of DAFs showed that the DAFs in LM were significantly different from those in SM, and there were also significant differences among LM, LD, and LY (Figure 3e). The DAMs were mapped to the KEGG biosynthetic pathway, and only 42 DAFs were annotated in the flavonoid biosynthetic pathway (Ko00941, Ko00942, Ko00943, and Ko00944). There were 19 DAFs annotated on the "flavonoid

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

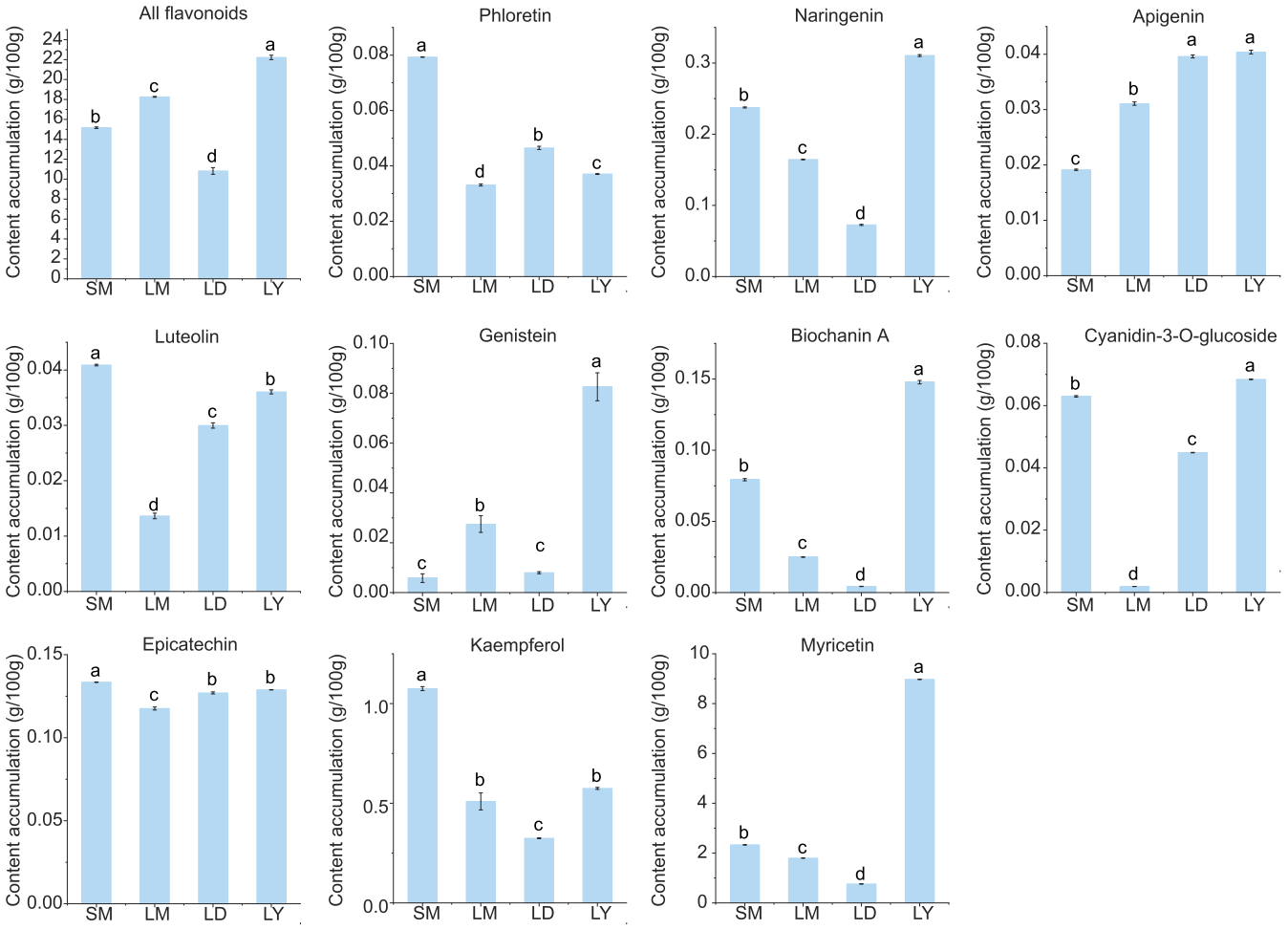


Figure 4. Content of flavonoid metabolites in different organs and at different growth and development stages of *L. polystachyus*. Different lowercase letters represent significant differences ($p < 0.05$).

biosynthesis" pathway (Ko00941). Among them, there were 11 DAFs in the comparison groups of different organs and 15 DAFs in the comparison groups at different growth and development stages (Table S2). Cluster analysis of 19 DAFs showed that in different organ comparison groups, except C01477 (apigenin), C09727 (epicatechin), and C90980 (neohesperidin), the other DAFs had higher accumulation in SM. In the comparison groups at different growth and development stages, C12127 (gallocatechin), C10107 (myricetin), C01477 (apigenin), C01514 (luteolin), C10192 (tricetin), C01617 (dihydroquercetin), C01460 (phlorizin chalcone), and C05903 (kaempferol) had higher accumulation in LY (Figure 3f).

Determination of Flavonoid Metabolites of *L. polystachyus*. By analyzing the UPLC-MS/MS results, we found that the content of various flavonoids differed significantly in different organs and at different growth and development stages in *L. polystachyus*. Various flavonoids have the same premetabolic pathway, and they compete with each other for the same substrates. To fully understand whether there are differences in the total flavonoid metabolism between *L. polystachyus*, the total flavonoid content was determined. The total flavonoid content of SM, LM, LD, and LY samples was determined using the total flavonoid content detection kit (Figure 4). The results showed that the total flavonoid content of LY was the highest (22.22 ± 0.081 g/100 g) and that of LD was the lowest (10.83 ± 0.330 g/100 g). The total flavonoid content of LM was 15.16 ± 0.081 g/100 g while that of SM was 18.26 ± 0.040 g/100 g.

After the basic analysis of the metabolites in the leaves and stems of *L. polystachyus*, the main flavonoid metabolites in the samples were quantitatively analyzed by UPLC (Figure 4). The experimental results were basically consistent with the metabolome data. In different organ comparison groups, naringenin, biochanin A, cyanidin-3-*O*-glucoside, epicatechin, and myricetin had higher accumulation in SM. In the comparison groups at different growth and development stages, apigenin, luteolin, kaempferol, and myricetin had higher accumulation in LY.

Transcriptome Data Analysis of *L. polystachyus*. After filtering the transcripts spliced by Trinity, 314,789 transcripts and 258,181 unigenes were obtained. The annotation of unigenes was based on KEGG, NR, Swiss-Prot, GO, COG/KOG, Trembl, and Pfam databases. A total of 258,181 unigenes were annotated, of which KOG annotation information was the least (88832, 34.41%) and NR annotation information was the most (176064, 68.19%) (Figure 5a). Compared with the NR database, it was found that the unigene sequence of *L. polystachyus* had the highest match with that of *Quercus suber* (76.46%), followed by *Juglans regia* (6.92%), *Vitis vinifera* (1.29%), and *Ziziphus jujuba* (0.62%) (Figure 5b). With the GO database analysis, it was found that 258,181

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

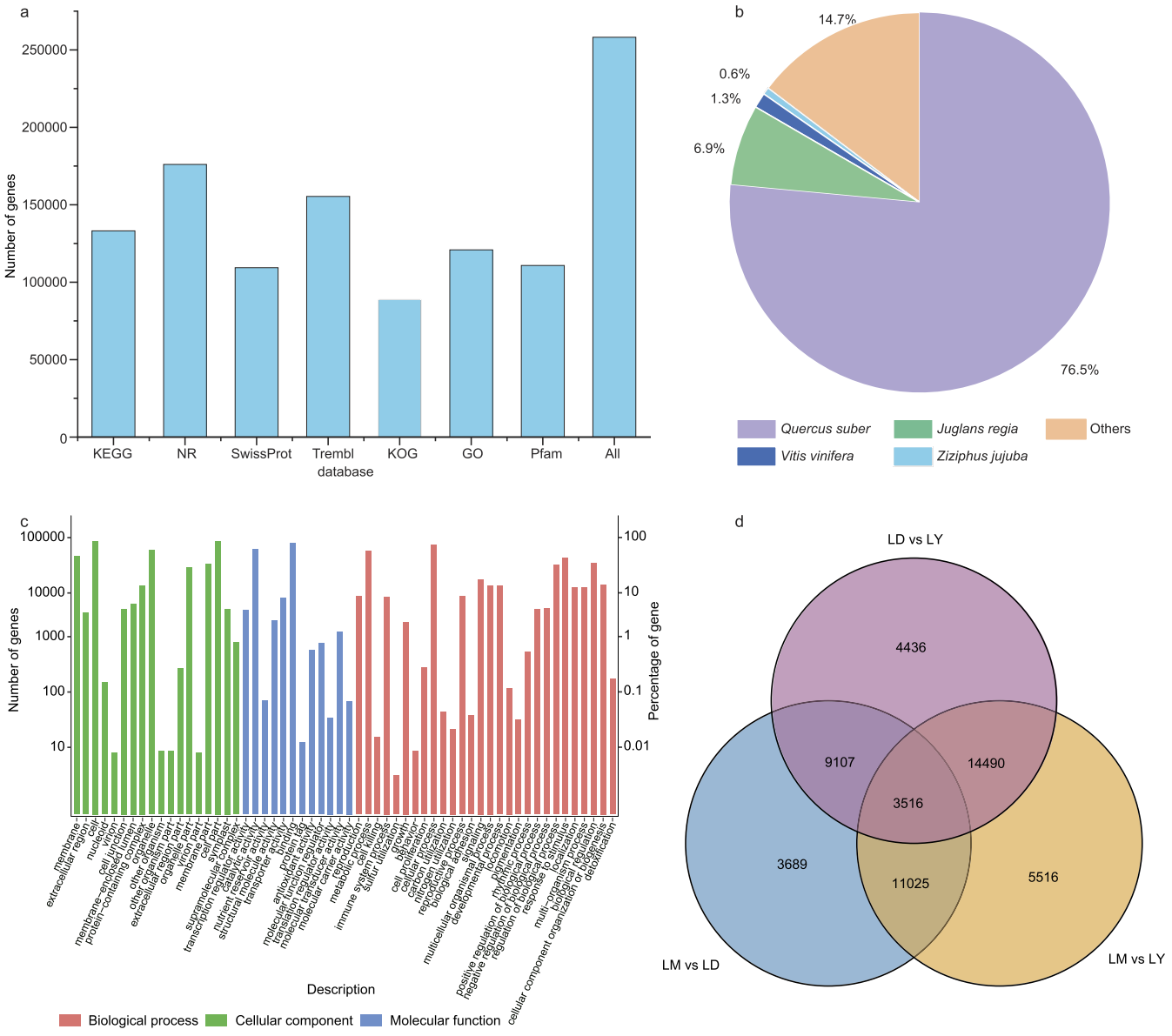


Figure 5. Unigene annotation and Venn diagram of DEGs at different growth and development stages of *L. polystachyus*. (a) Number of unigenes in seven data annotation information. (b) Unigene annotation information in the Nr database. (c) Unigene annotation information in the GO database. (d) Venn diagram of DEGs at different growth and development stages.

unigenes were classified into 59 functional groups, and these groups were divided into 3 categories: cellular component group mainly involving cells (86997, 33.70%), cell parts (86854, 33.64%), and organelles (59925, 23.21%), the molecular function group mainly involving binding (79262, 30.70%) and catalytic activity (62108, 24.06%), and the biological process group mainly involving the cellular process (73071, 28.30%) and metabolic process (57410, 22.24%) (Figure 5c).

The criteria of FDR < 0.01 and $|\log_2 FC| \geq 1$ were used for screening significant differences in the expression of genes, and DEGs were screened in transcriptome data. In the "LM versus SM" group, 23,327 DEGs were detected, including 11,590 upregulated DEGs and 11,737 downregulated DEGs. In the "LM versus LD" group, 27,337 DEGs were detected, including 13,330 upregulated DEGs and 14,007 downregulated DEGs. In the "LM versus LY" group, 34,547 DEGs were detected, including 17,393 upregulated DEGs and 17,154 downregulated DEGs. In the "LD versus LY" group, 31,549 DEGs were detected, including 16,233 upregulated DEGs and 15,316 downregulated DEGs. The Venn diagram results showed that a total of 51,779 DEGs were detected in the comparison groups of "LM versus LD," "LM versus LY," and "LD versus LY", among which 3516 DEGs were common. A total of 3689 DEGs were differentially expressed only in the "LM versus LD" group, 5516 DEGs were differentially expressed only in the "LM versus LY" group, and 4436 DEGs were differentially expressed only in the "LD versus LY" group (Figure 5d).

All DEGs were annotated to the KEGG pathway for enrichment analysis. The results showed that they were primarily enriched in the "biosynthesis of secondary metabolites", "flavonoid biosynthesis", "isoflavonoid biosynthesis", "MAPK signaling pathway plant", "monobactam biosynthesis", "plant hormone signal transduction", "plant−pathogen interaction", and "zeatin biosynthesis" of all comparison groups. They were primarily enriched in the "ABC transporters" and

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

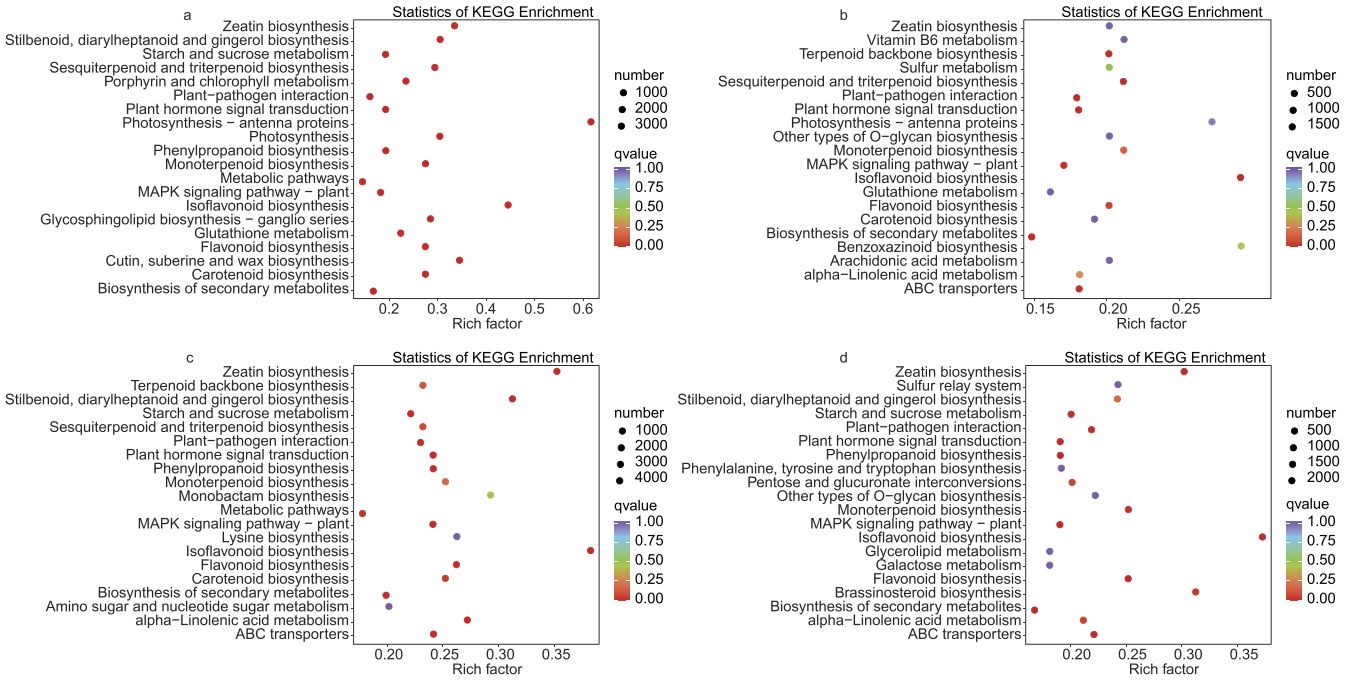


Figure 6. Enrichment analysis of DEGs in the KEGG pathway. (a) KEGG enrichment of DEGs in the "LM versus SM" group. (b) KEGG enrichment of DEGs in the "LM versus LD" group. (c) KEGG enrichment of DEGs in the "LM versus LY" group. (d) KEGG enrichment of DEGs in the "LD versus LY" group.

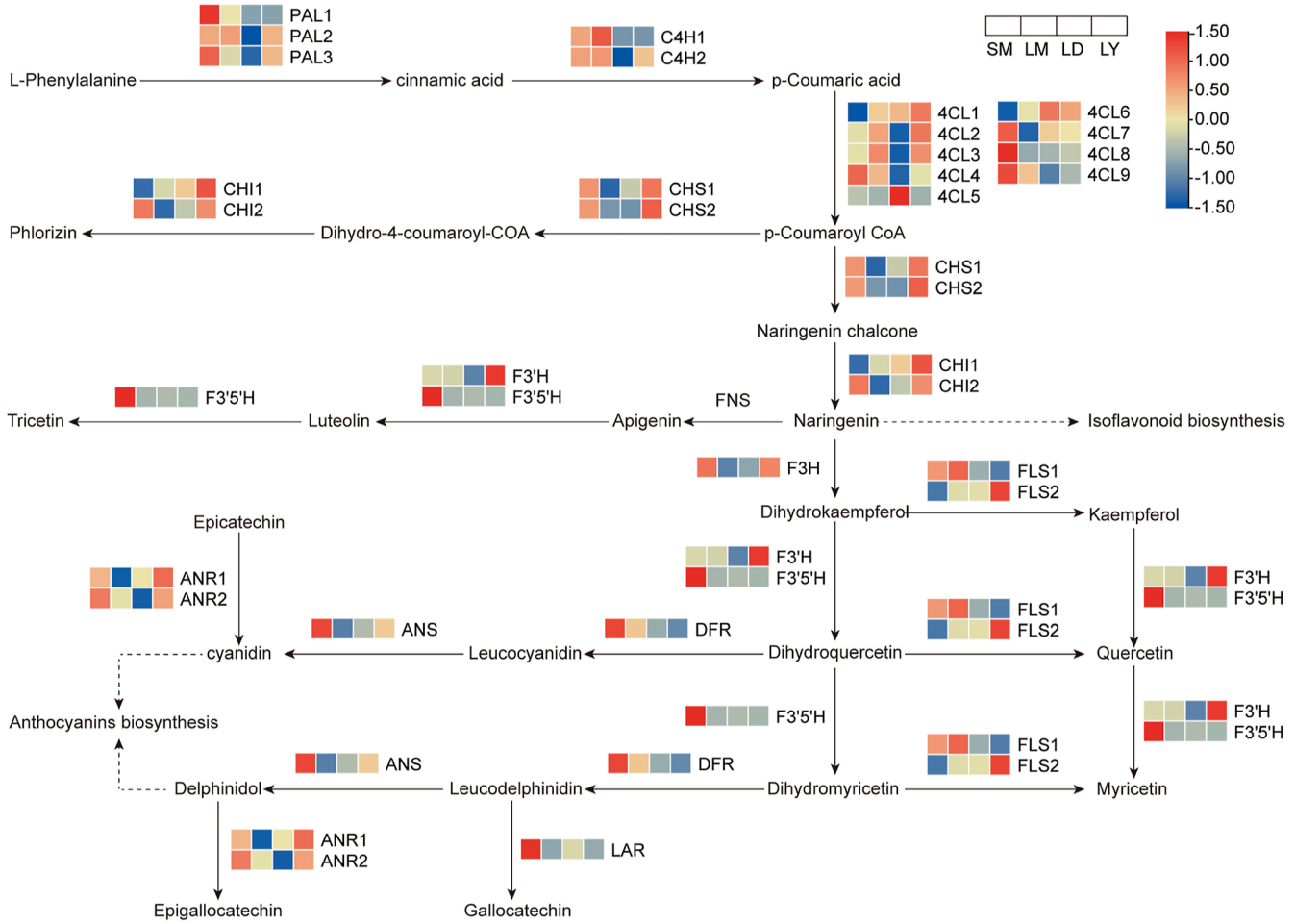


Figure 7. Flavonoid biosynthetic pathway of *L. polystachyus* and cluster heat map of the FPKM value of DEGs.

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

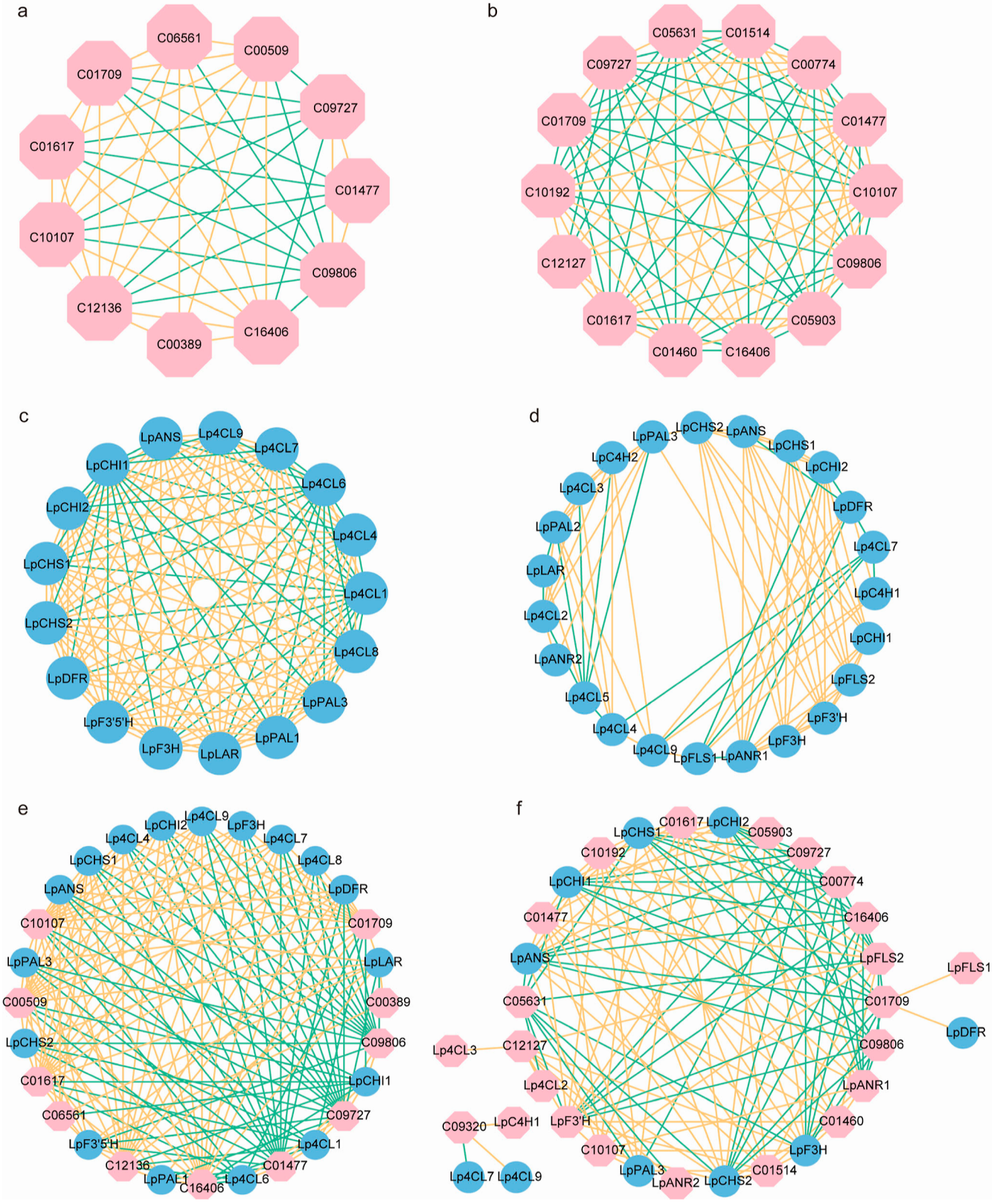


Figure 8. Network diagram of the correlation between DAFs and DEGs in the flavonoid biosynthetic pathway of *L. polystachyus*. (a) Correlation network diagram between DAFs in different organ comparison groups. (b) Correlation network diagram between DAFs at different growth and development stage comparison groups. (c) Correlation network diagram between DEGs in different organ comparison groups. (d) Correlation network diagram between DEGs in the comparison groups at different growth and development stages. (e) Correlation network diagram between DAFs and DEGs in different organ comparison groups. (f) Correlation network diagram between DAFs and DEGs at different growth and development stage comparison groups. Pink octagons: DAFs, blue rounds: DEGs; orange lines: positive correlations, green lines: negative correlations.

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

"α-linolenic acid metabolism" of each growth and development stage (Figure 6a−d).

Expression Analysis of Flavonoid Biosynthesis-Related DEGs in *L. polystachyus*. The metabolism of flavonoids involves the phenylpropane biosynthesis pathway (Ko00940) and the flavonoid biosynthetic pathway (Ko00941, Ko00942, Ko00943, and Ko00944). By referring to the relevant metabolic pathways in the KEGG database and related literature,[17] we speculated on the process of flavonoid biosynthesis in *L. polystachyus* (Figure 7). According to the enrichment results of DEGs in the KEGG pathway, 28 DEGs were identified in the flavonoid biosynthetic pathway, including phenylalanine ammonia-lyase (*LpPAL1*−*LpPAL3*), *trans*-cinnamate 4-monooxygenase (*LpC4H1* and *LpC4H2*), 4-coumarate-CoA ligase (*Lp4CL1*−*Lp4CL9*), chalcone synthase (*LpCHS1* and *LpCHS2*), chalcone isomerase (*LpCHI1* and *LpCHI2*), naringenin 3-dioxygenase (*LpF3H*), flavonoid 3′-monooxygenase (*LpF3′H*), flavonoid 3′,5′-hydroxylase (*LpF3′5′H*), flavonol synthase (*LpFLS1* and *LpFLS2*), dihydroflavonol 4-reductase (*LpDFR*), anthocyanidin synthase (*LpANS*), anthocyanidin reductase (*LpANR1* and *LpANR2*), and leucoanthocyanidin reductase (*LpLAR*) (Table S3). There were 17 DEGs in different organ comparison groups and 24 DEGs in comparison groups at different growth and development stages, of which 12 DEGs were common differences (*LpPAL3*, *Lp4CL6*, *Lp4CL7*, *Lp4CL9*, *LpCHS1*, *LpCHS2*, *LpCHI1*, *LpCHI2*, *LpF3H*, *LpDFR*, *LpLAR*, and *LpANS*). Clustering analysis of 28 DEGs related to flavonoid biosynthesis showed that in different organ comparison groups, except *Lp4CL1*, *Lp4CL6*, and *LpCHI1*, the expression levels of the other DEGs were relatively high in SM. In the comparison groups at different growth and development stages, *Lp4CL4*, *Lp4CL9*, *LpC4H1*, *LpDFR*, and *LpFLS1* were found to be highly expressed in LM, and *LpPAL3*, *Lp4CL2*, *LpCHS1*, *LpCHS2*, *LpCHI1*, *LpCHI2*, *LpF3H*, *LpF3′H*, *LpFLS2*, *LpANS*, *LpANR1*, and *LpANR2* were highly expressed in LY (Figure 7).

Correlation Analysis of DAFs and DEGs in *L. polystachyus*. There were significant differences between DAMs in different organs and at different growth and development stages in *L. polystachyus*. Studies were performed to further understand whether there is a correlation of content accumulation between DAFs, between DEGs, between DAFs and DEGs in different organs, and at different growth and development stages of *L. polystachyus* and to determine the key genes during flavonoid biosynthesis in *L. polystachyus* (PCC ≥ 0.8, $p < 0.01$).

Correlation analysis between the DAFs annotated in the Ko00941 biosynthesis pathway of *L. polystachyus* (PCC ≥ 0.8, $p < 0.01$) was also carried out. Among the comparison groups of different organs, C12136 (epigallocatechin) had the least linear correlation with 10 metabolites and showed a negative correlation with C09806 (neohesperidin), C09727 (epicatechin), C01477 (apigenin), and positive correlation with 7 metabolites. C00389 (quercetin) had the least linear correlation with four metabolites and showed a positive correlation with C06561 (naringin chalcone), C00509 (naringin), C09727 (epicatechin), and C16406 (phlorizin chalcone). There was a positive correlation between C06561 (naringin chalcone) and C00509 (naringin), between C00509 (naringin) and C01617 (dihydroquercetin), and between C01617 (dihydroquercetin) and C10107 (myricetin) (Figure 8a). In comparison groups at different growth and development stages, there was a positive correlation between C01477 (apigenin), C10192 (tricetin), and C01514 (luteolin). C01617 (dihydroquercetin) was positively correlated with C09727 (epicatechin), C12136 (epigallocatechin), C10107 (myricetin), and C05903 (kaempferol), while C05903 (kaempferol) was positively correlated with C10107 (myricetin) (Figure 8b).

We performed a correlation analysis of 28 DEGs in the flavonoid biosynthetic pathway of *L. polystachyus* (PCC ≥ 0.8, $p < 0.01$). In the comparison groups of different organs, *Lp4CL1* showed a linear correlation with 16 DEGs, a positive correlation with *LpCHI1* and *Lp4CL6*, and a negative correlation with the other DEGs. Both *LpCHI1* and *Lp4CL6* were only positively correlated with *Lp4CL1* and negatively correlated with the other DEGs. *LpCHS1*, *LpCHS2*, *LpCHI2*, *LpF3H*, *LpF3′5′H*, *LpDFR*, *LpANS*, and *LpLAR* had a positive correlation with each other (Figure 8c). *LpCHI2* and *LpANR*1 had the most significant linear relationship among the comparison groups at different growth and development stages, and both of them were associated with 10 DEGs. *LpANR2* had the least significant linear relationship and positively correlated with *LpPAL3* and *Lp4CL2*. *LpCHS1*, *LpCHS2*, *LpCHI1*, *LpCHI2*, *LpF3H*, *LpF3′H*, *LpFLS2*, *LpANS*, and *LpANR1* had a positive correlation with each other (Figure 8d).

To explore the relationship between DAFs and DEGs during flavonoid biosynthesis, the correlation between DEGs and DAFs was calculated (PCC ≥ 0.8, $p < 0.01$). In the comparison groups of different organs, *LpCHS1*, *LpCHS2*, and *LpCHI2* were positively correlated with C06561 (naringin chalcone) and C00509 (naringin). *LpF3H* was positively correlated with C00509 (naringin) and C01617 (dihydroquercetin), and *LpF3′5′H* was positively correlated with C10107 (myricetin) and C00389 (quercetin). *LpDFR*, *LpANS*, and *LpLAR* were positively correlated with C01617 (dihydroquercetin) and C12136 (epigallocatechin) (Figure 8e). In the comparison groups at different growth and development stages, there were negative correlations between C00774 (phloretin), C05631 (eriodictyol), C09806 (neohesperidin), C09727 (epicatechin), and DEGs. *LpPAL3*, *LpCHS1*, *LpCHS2*, *LpCHI1*, *LpCHI2*, and *LpF3′H* were positively correlated with C01477 (apigenin), C01514 (luteolin), and C10192 (tricetin). *LpPAL3*, *LpCHS1*, *LpCHS2*, *LpCHI1*, *LpCHI2*, and *LpF3H* were positively correlated with C01617 (dihydroquercetin), and *LpFLS2* was positively correlated with C01617 (dihydroquercetin) and C10107 (myricetin). *LpF3′H* was positively correlated with C05903 (kaempferol) and C10107 (myricetin) (Figure 8f).

Key DEGs and DAMs in the flavonoid biosynthesis of *L. polystachyus* have a consistent expression pattern, and there is a significantly positive correlation between key structural genes and upstream and downstream metabolites. The contents of DAFs, including naringenin chalcone, naringenin, dihydroquercetin, epigallocatechin, quercetin, and myricetin in SM were relatively high in different organ comparison groups, and there was a significant positive correlation between DAFs. The expression patterns of flavonoid-related DEGs, including *LpPAL1*, *LpPAL3*, *Lp4CL4*, *Lp4CL7*, *Lp4CL8*, *Lp4CL9*, *LpCHS1*, *LpCHS2*, *LpCHI2*, *LpF3H*, *LpF3′5′H*, *LpDFR*, and *LpANS* were consistent with the DAFs, and flavonoid-related DEGs have relatively high expressions in SM. There was a significant positive correlation between DEGs and between the DAFs and their upstream and downstream DEGs. The results of the comprehensive analysis of DAFs and DEGs in different

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

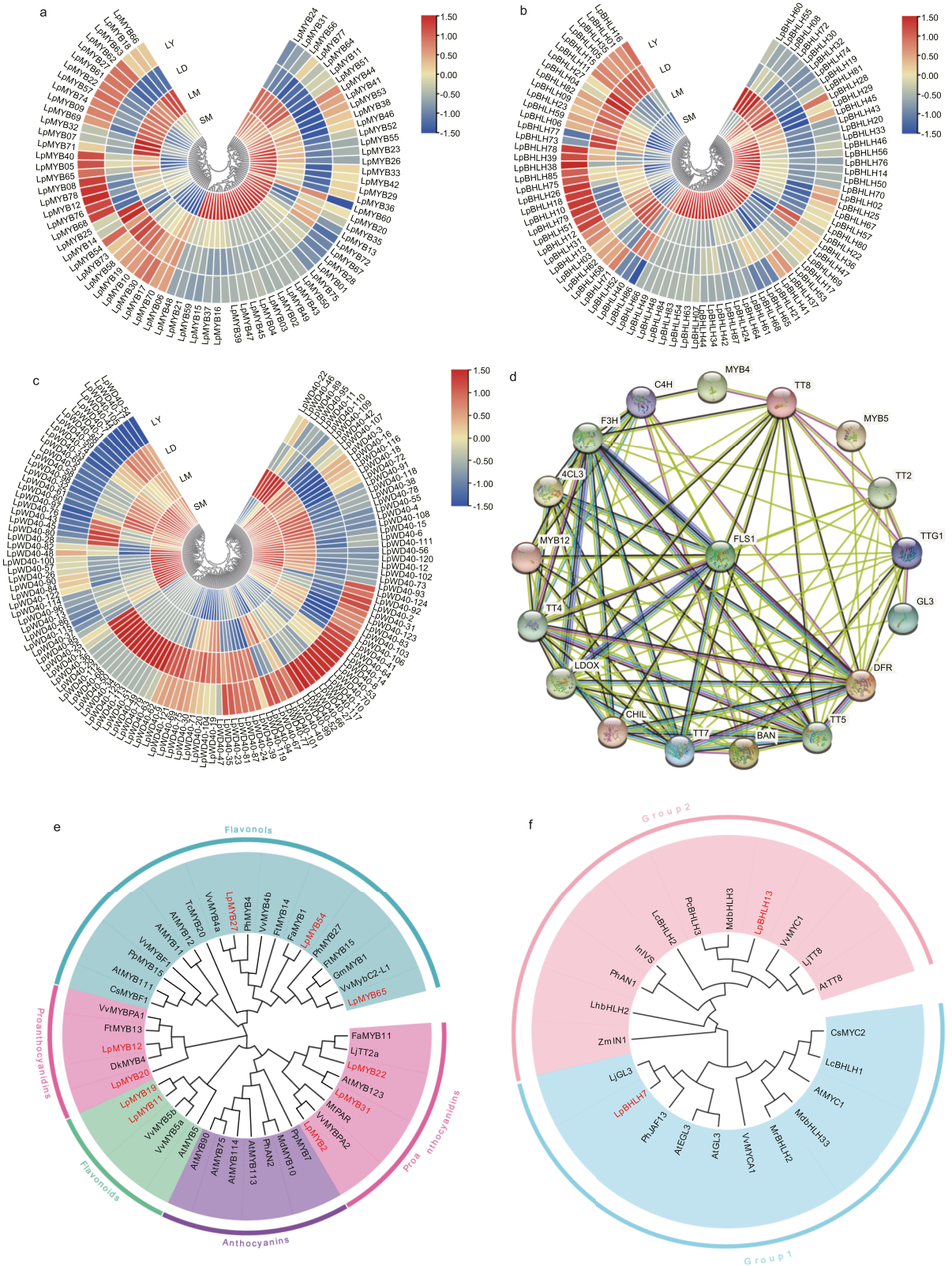


Figure 9. Analysis of DTFs in *L. polystachyus*. (a) Cluster analysis of LpMYBs. (b) Cluster analysis of LpbHLHs. (c) Cluster analysis of LpWD40s. (d) Protein−protein interaction network of DTFs and DEGs. Deep-sky-blue line: from curated databases, violet−red line: experimentally determined, green line: gene neighborhood, red line: gene fusions, blue line: gene co-occurrence, olive drab line: text mining, purple line: coexpression, black line: protein homology. (e) Phylogenetic tree of MYBs. (f) Phylogenetic tree of bHLHs.

organ comparison groups of *L. polystachyus* showed that *LpPAL1*, *LpPAL3*, *Lp4CL4*, *Lp4CL7*, *Lp4CL8*, *Lp4CL9*, *LpCHS1*, *LpCHS2*, *LpCHI2*, *LpF3H*, *LpF3′5′H*, *LpDFR*, and *LpANS* play a key role in the accumulation of flavonoids in *L. polystachyus*. In the comparison groups at different growth and development stages of *L. polystachyus*, the expression patterns

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

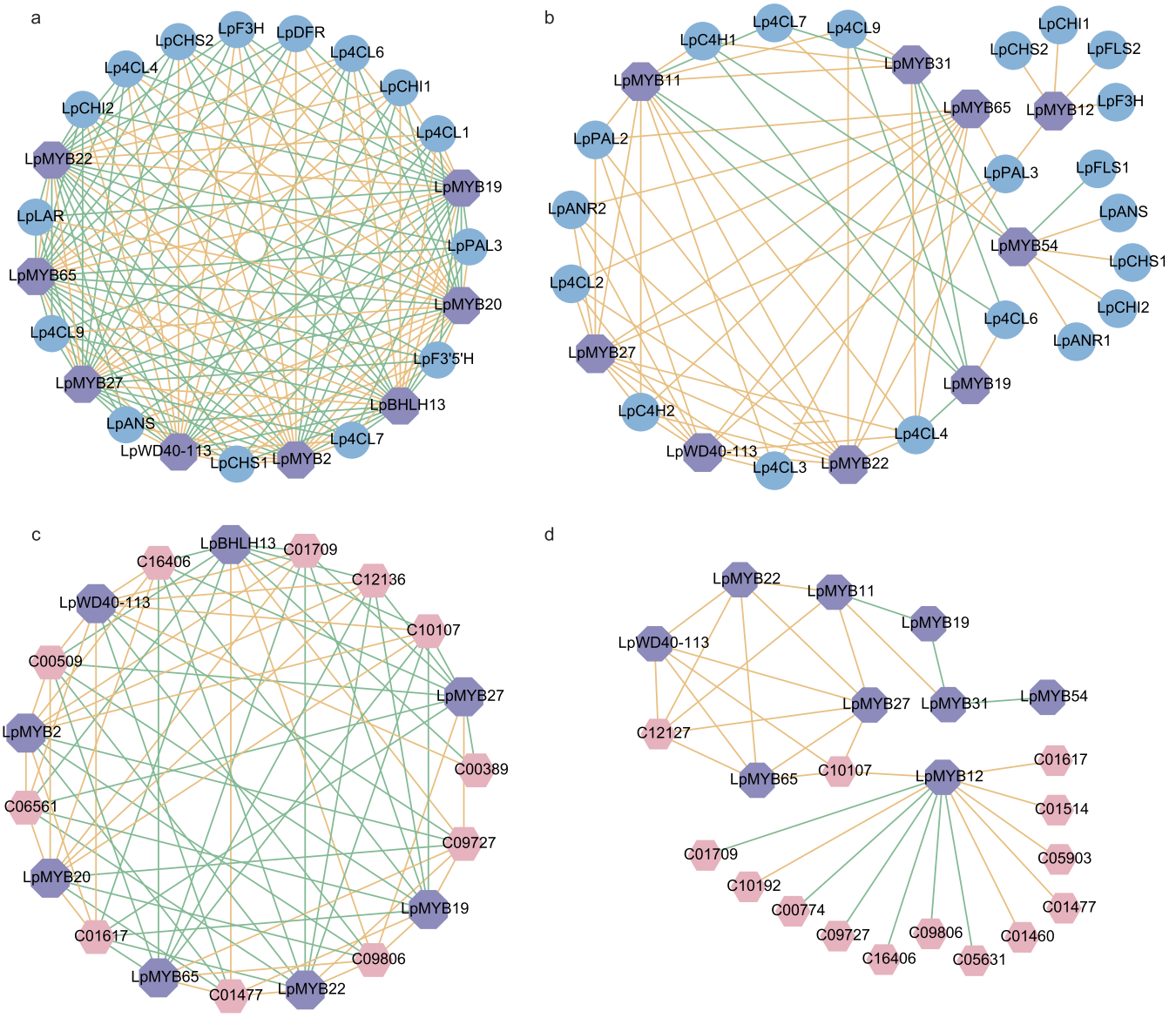


Figure 10. Network diagram of the correlation between DTFs, DEGs, and DAMs in the flavonoid biosynthetic pathway of *L. polystachyus*. (a) Correlation network diagram between DTFs and DEGs in different organ comparison groups. (b) Correlation network diagram between DTFs and DEGs at different growth and development stage comparison groups. (c) Correlation network diagram between DTFs and DAMs in different organ comparison groups. (d) Correlation network diagram between DTFs and DAMs in the comparison groups at different growth and development stages. Pink octagons: DAFs, blue rounds: DEGs, purple hexagons: DTFs; orange lines: positive correlations, green lines: negative correlations.

of apigenin, luteolin, tricetin, dihydroquercetin, gallocatechin, kaempferol, and myricetin, associated with *LpPAL3*, *Lp4CL2*, *Lp4CL7*, *LpCHS1*, *LpCHS2*, *LpCHI1*, *LpCHI2*, *LpF3H*, *LpF3′H*, *LpFLS2*, *LpANS*, *LpANR1*, and *LpANR2* were consistent, and their expression levels in LY were relatively high. There were significantly positive correlations between apigenin, luteolin, and tricetin, between dihydroquercetin and gallocatechin, and between dihydroquercetin, kaempferol, and myricetin at different growth and development stages of *L. polystachyus*. Upstream and downstream DEGs were significantly and positively correlated and so were DAFs and their upstream and downstream DEGs. The comprehensive analysis of DAFs and DEGs in the comparison groups at different growth and development stages of *L. polystachyus* showed that *LpPAL3*, *Lp4CL2*, *LpCHS1*, *LpCHS2*, *LpCHI1*, *LpCHI2*, *LpF3H*, *LpF3′H*, *LpFLS2*, *LpANS*, *LpANR1*, and *LpANR2* play a key role in the accumulation of flavonoids in *L. polystachyus* and verified the correctness of the flavonoid biosynthetic pathway of *L. polystachyus* (Figure 7). Phenylalanine was catalyzed by LpPAL3, Lp4CL7, and LpC4H to produce coumaroyl-CoA that entered the flavonoid biosynthetic pathway and was catalyzed by LpCHS1, LpCHS2, and LpCHI2 to produce naringin chalcone and naringin. Naringin generates a series of flavone metabolites under the catalytic action of LpF3′H and LpF3′5′H and dihydroflavonols catalyzed by LpF3H, LpF3′H, and LpF3′5′H. Dihydroflavonols were catalyzed by LpFLS2 to form flavonols and by LpDFR, LpANS, LpANR1, and LpANR2 to produce flavanones and anthocyanins.

Analysis of DTFs of *L. polystachyus*. Transcriptional expression of DEGs in the flavonoid biosynthetic pathway of *Arabidopsis* was regulated by multiple TFs.[18] R2R3-MYB,

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

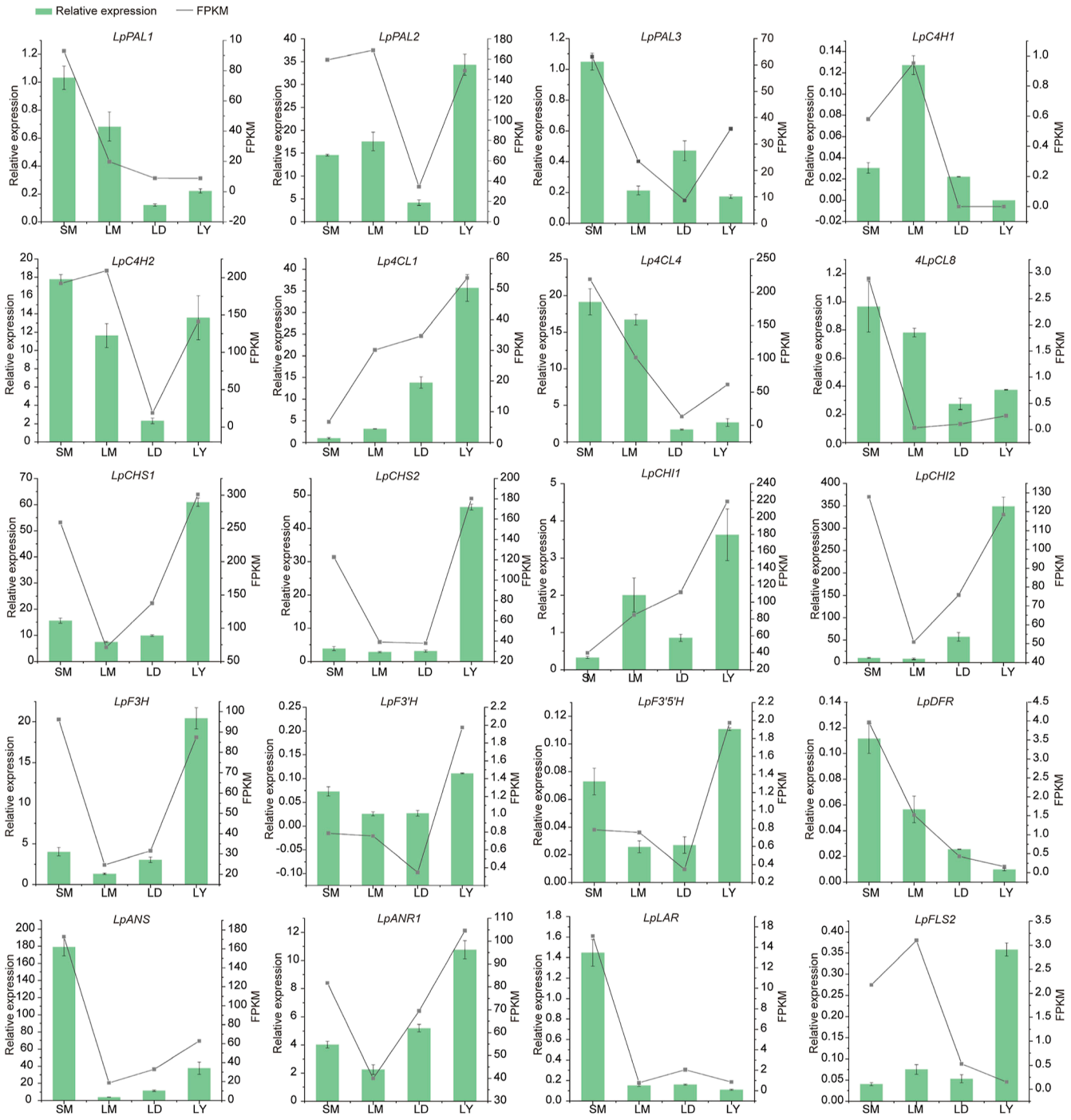


Figure 11. Analysis of the relative expression of structural genes in the flavonoid biosynthetic pathway of *L. polystachyus* by qRT-PCR.

bHLH, and WD40 could act alone or in concert with others to control multiple enzymatic steps in the flavonoid biosynthetic pathway of different species. To explain the regulatory background of the DAMs in *L. polystachyus*, we studied the expression profiles of *LpMYB*, *LpbHLH*, and *LpWD40* DTFs in different organ comparison groups and the comparison groups at different growth and development stages. A total of 78 *LpMYBs*, 87 *LpbHLHs*, and 125 *LpWD40s* were screened from the DEGs of *L. polystachyus* for cluster analysis. The results showed that in the SM of different organ comparison groups, the expression of 28 *LpMYB*, 34 *LpbHLH*, and 15 *LpWD40* DTFs was high while those of 18 *LpMYB*, 16 *LpbHLH*, and 17 *LpWD40* DTFs were low. In the comparison groups at different growth and development stages, 20 *LpMYB*, 15 *LpbHLHs*, and 54 *LpWD40* DTFs were highly expressed in LM while 14 *LpMYBs* 13 *LpbHLHs*, and 41 *LpWD40* DTFs in LY (Figure 9a−c).

In order to screen out TFs associated with flavonoid accumulation in *L. polystachyus*, STRING 11 was used to reconstruct a protein−protein interaction network of DTFs and DEGs related to flavonoid biosynthesis using homologous protein data of *Arabidopsis thaliana* (combined score > 0.7, high confidence) (Table S4). The results of protein−protein interaction showed that 10 LpMYBs (LpMYB2, LpMYB11, LpMYB12, LpMYB19, LpMYB20, LpMYB22, LpMYB27, LpMYB31, LpMYB54, and LpMYB65), 2 LpbHLHs

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

(LpbHLH7 and LpbHLH13) DTFs, and 1 LpWD40 (LpWD40-113) were predicted to be involved in the flavonoid biosynthesis of *L. polystachyus*. *TT2* (*LpMYB22*), *TT8* (*LpBHLH13*), and *TTG1* (*LpWD40-113*) form MBW complexes that activate the transcription of *TT5* (*LpCHI1*), *F3H* (*LpF3H*), *TT7* (*LpF3′H*, *LpF3′5′H*), *FLS1* (*LpFLS1*, *LpFLS2*), *DFR* (*LpDFR*), *LDOX* (*LpANS*), and *BAN* (*LpANR1* and *LpANR2*) genes and regulate flavonol and anthocyanin biosynthesis (Figure 9d).

To understand the functions of LpMYBs and LpbHLHs in regulating flavonoid biosynthesis of *L. polystachyus*, we constructed a phylogenetic tree by combining the above-predicted LpMYBs and LpbHLHs DTFs with functional MYBs and bHLHs regulating the flavonoid biosynthesis of other species by using MEGA-X (Figure 9e,f). According to the classification method of MrMYBs in the flavonoid biosynthesis of *Myrica rubra* by Cao et al.,[19] we divided the MYBs of the phylogenetic tree into five groups. Among them, LpMYB2, LpMYB22, and LpMYB31 were closely clustered with AtMYB123 (AtTT2) and VvMYBPA2, which regulated PA production, and LpMYB12 and LpMYB20 were clustered with VvMYBPA1, which was involved in PA generation. LpMYB19 and LpMYB11 clustered together with MYB related to flavonoid biosynthesis, while LpMYB27, LpMYB54, and LpMYB65 clustered together with MYB related to flavanol biosynthesis. AtPAP1 (AtMYB75) is a well-known gene of *Arabidopsis* regulating anthocyanin biosynthesis, and there is no LpMYB that clusters with AtPAP1. The absence of LpMYB TFs is associated with anthocyanin synthesis, which might be because of the low number of anthocyanin-like species in *L. polystachyus*. The difference in anthocyanin metabolites was only cyanidin-3-*O*-glucoside, and no related DTFs were involved in anthocyanin biosynthesis. We divided LpbHLH TFs into two groups, in which LpbHLH13 clustered with PhAN1, LcBHLH2, and AtTT8, while LpbHLH7 clustered with PhJAF13, AtEGL, and LcBHLH2.

To further confirm the regulatory mechanisms of DTFs in the flavonoid biosynthesis of *L. polystachyus*, we analyzed the correlation between DTFs and DEGs. In the different organ comparison groups, *LpBHLH13*, *LpMYB19*, *LpMYB22*, *LpMYB27*, and *LpMYB65* showed positive correlations with *Lp4CL1*, *LpCHI1*, and *Lp4CL6*, negative correlations with other genes, and a positive correlation with each other. *LpMYB2*, *LpMYB20*, *LpbHLH7*, and *LpWD40-113* were positively correlated with *LpCHS1*, *LpCHS2*, *LpCHI2*, *LpF3H*, *LpF3′5′H*, *LpDFR*, *LpANS*, and *LpLAR* (Figure 10a). In the comparison groups at different developmental stages, *LpMYB12* showed positive correlations with *LpCHS2*, *LpCHI1*, *LpF3H*, *LpF3′H*, and *LpFLS2*. *LpMYB54* showed positive correlations with *LpCHS1*, *LpCHI2*, *LpANS*, and LpANR1. *LpMYB22*, *LpMYB27*, *LpMYB65*, *LpMYB11*, and *LpWD40-113* were positively correlated with *LpC4H2*, *Lp4CL2*, *Lp4CL4*, and *Lp4CL9*. *LpMYB22*, *LpMYB27*, and *LpMYB65* were positively correlated with *LpANR2*. *LpMYB31* was positively correlated with *LpC4H1*, *Lp4CL9*, *Lp4CL4*, and *LpMYB11* (Figure 10b).

To explore the relationship between DTFs and DAFs during flavonoid biosynthesis, the correlations between them were calculated (PCC $\geq$ 0.8, $p < 0.01$). In the different organ comparison groups, *LpBHLH13*, *LpMYB19*, *LpMYB22*, *LpMYB27*, and *LpMYB65* showed positive correlations with C09806 (neohesperidin), C09727 (epicatechin), and C01477 (apigenin) and negative correlations with other metabolites. *LpMYB2*, *LpMYB20*, and *LpWD40-113* showed positive correlations with C06561 (naringin chalcone), C00509 (naringin), C01617 (dihydroquercetin), C12136 (epigallocatechin), and C10107 (myricetin). *LpWD40-113* showed positive correlations with C00389 (quercetin) (Figure 10c). In comparison groups at different growth and development stages, *LpMYB12* showed positive correlations with C01477 (apigenin), C10107 (myricetin), C01514 (luteolin), C10192 (tricetin), and C01617 (dihydroquercetin). *LpMYB22*, *LpMYB27*, *LpMYB65*, and *LpWD40-113* showed positive correlations with C12127 (gallocatechin) and C10107 (myricetin) (Figure 10d).

qRT-PCR Experiment Verification. A total of 20 DEGs were randomly selected for qRT-PCR analysis, and the results showed that the relative expression patterns of DEGs were similar to those of transcriptome sequencing data. In the comparison groups of different organs, the expression levels of *LpPAL1*, *Lp4CL2*, *LpCHS1*, *LpCHS2*, and *LpCHI2* in SM were relatively high. In the comparison groups at different growth and development stages, *LpPAL2*, *Lp4CL9*, *LpC4H1*, *LpDFR*, and *LpFLS2* were highly expressed in LM, and *LpPAL3*, *LpCHS1*, *LpCHS2*, *LpCHI1*, *LpCHI2*, *LpF3H*, *LpF3′H*, *LpANS*, and *LpANR1* were highly expressed in LY (Figure 11).

DISCUSSION

In this study, we performed metabolomic and transcriptomic analyses on samples of different organs and at different growth and development stages of *L. polystachyus*. The results showed that the total flavonoid content of SM was higher than that of LM, and the content of *L. polystachyus* was highest in young leaves and decreased in mature leaves.[13] The flavonol content is mostly the highest in the early stage and gradually decreases with the growth and development of *L. polystachyus*. Flavonoids mainly include kaempferol, myricetin, quercetin, and their derivatives. Among them, afzelin is the flavonol compound with the highest accumulation in *L. polystachyus* and is a kaempferol derivative. Naringenin chalcone is the most important chalcone compound and is highly accumulated in SM. Dihydroflavonoids include naringenin and sageol, and the former is a precursor substance for the synthesis of other flavonoids.[20] Compared to that of LM, naringin had higher accumulation in SM and LY. Cyanidin-3-*O*-glucoside is the only differential anthocyanin compound with higher accumulation in SM. The content of most flavone compounds is similar in different organs and gradually decreases with the growth and development of *L. polystachyus*, such as apigenin, luteolin, and tricetin. Catechins and epicatechins are the main flavanol compounds in *L. polystachyus*, and epicatechins have higher accumulation in SM. Most isoflavones gradually decrease with the growth and development of *L. polystachyus*, and prunetin and sissotrin show higher accumulation in LY. There are a few types of anthocyanins identified in *L. polystachyus*, which may be because of the absence of organ tissues requiring pigment accumulation in *L. polystachyus*. The leaf development period of *L. polystachyus* is identified according to the leaf extension instead of the leaf color shade. It has been shown that flavonoid biosynthesis of plants is a complex network of the regulatory process where various genes and enzymes play a regulatory role. According to the consistent expression pattern of DEGs and DAFs and the significant correlation between DEGs and their upstream and downstream metabolites, we speculated the key genes in the flavonoid biosynthesis process of *L. polystachyus*, including

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

LpPAL3, *Lp4CL7*, *LpCHS1*, *LpCHS2*, *LpCHI2*, and *LpF3H*. As PAL, C4H, and 4CL were involved in flavonoid biosynthetic pathways and lignin biosynthesis, the correlation patterns of *LpPAL* and *Lp4CL* expression are slightly different from the other structural genes of flavonoid biosynthesis of *L. polystachyus*. The contents of liquiritigenin, isoliquiritigenin, isoliquiritin, and total flavonoids in transgenic hairy roots with overexpressing *GuCHS* were significantly higher than those in the wild-type hairy roots in *Glycyrrhiza uralensis*.[21] Expression of *OjCHI* in *Arabidopsis* tt5 mutant restored the accumulation of anthocyanins and flavonols.[22] The silencing of *FaF3H* in strawberries significantly decreased the contents of flavonols and anthocyanins.[23] Functional verification experiments have confirmed the role of CHI, CHS, and F3H in flavonoid biosynthesis, which is similar to our results and indicates that the mechanism of flavonoid biosynthesis of different plants is relatively conservative.

MYB, bHLH, and WD40 regulate flavonoid biosynthesis by activating or inhibiting the expression of structural genes. MYB is the most important in the plant flavonoid pathway. After it binds to specific DNA regulatory elements in the promoter region of the target gene, transcriptional activation is initiated.[24] PA and anthocyanin-specific MYBs are also required to bind to bHLHs and WD40s repeat proteins to form MBW complexes to promote transcription. MYBs proteins act as direct activators of structural genes and activators of genes encoding bHLHs.[25] AtMYB111 can bind to specific *cis*-elements in *AtCHS*, *AtF3H*, and *AtFLS1* promoters to activate their transcription in *A. thaliana*.[26] Interaction of *VvMYC1* with *VvMYB5a* and *VvMYB5b* induces the initiation of genes involved in anthocyanin biosynthesis.[27] A total of 10 MYB, 2 BHLH, and 1 WD40 TFs were predicted to be associated with flavonoid synthesis. According to the clustering results, we speculated that LpMYB19 and LpMYB11 were involved in flavonoid biosynthesis; LpMYB12, LpMYB20, LpMYB22, LpMYB31, and LpMYB2 were involved in PA biosynthesis; and LpMYB27, LpMYB54, and LpMYB65 were involved in flavonol biosynthesis. The bHLH TFs involved in flavonoid biosynthesis have been divided into two major groups: bHLH2/AN1/TT8 and bHLH/JAF13/EGL3 clades. The former directly activates the expression of genes related to flavonoid biosynthesis, while the latter regulates the transcription of bHLH2/AN1/TT8.[28] LpbHLH13 belongs to the bHLH2/AN1/TT8 branch, while LpbHLH7 belongs to the bHLH1/JAF13/EGL3 branch that indirectly activates the expression of genes related to flavonoid biosynthesis. In addition to activating gene expressions, TFs acted as repressors to repress the expression of structural genes.[29] AtMYB4 inhibits flavonoid accumulation by repressing the expression of the gene encoding arogenate dehydratase 6 (ADT6).[30] The results of correlation analysis of DTFs and DEGs showed that in the comparison groups of different organs, the expressions of *LpBHLH13*, *LpMYB19*, *LpMYB22*, *LpMYB27*, and *LpMYB65* were significantly negatively correlated with DEGs related to flavonoid biosynthesis, and *LpMYB2*, *LpMYB20*, and *LpWD40-113* were positively correlated with DEGs. LpMYB27 and LpMYB65 clustered together with MYB repressors such as FaMYB1, FaMYB14, and VvMYBC2L-1, which inhibited gene expression.[31] In the comparison groups at different growth and development stages, *LpMYB54* and *LpMYB12* positively regulated the expression of key genes involved in flavonoid biosynthesis of *L. polystachyus*. As *LpMYB12* showed positive correlations with *LpCHS2*, *LpCHI1*, *LpF3H*, *LpF3′H*, and *LpFLS*, we speculated that LpMYB12 is mainly involved in flavonol biosynthesis. As *LpMYB54* showed positive correlations with *LpCHS1*, *LpCHI2*, *LpANS*, and *LpANR1*, we speculated that LpMYB54 is mainly involved in anthocyanins and flavanol biosynthesis. The correlation analysis between DTFs and DAMs showed that LpMYB2, LpMYB20, and LpWD40-113 were positively correlated with flavonoids in different organ comparison groups. LpMYB12 showed a positive correlation with flavonoids of the comparison groups at different growth and development stages. The results showed that LpMYB2, LpMYB20, LpMYB54, LpMYB12, and LpWD40-113 positively regulated the biosynthesis of flavonoids by regulating key genes involved in the flavonoid biosynthesis of *L. polystachyus*.

In summary, through the combined analysis of metabolomics and transcriptomics, we inferred the flavonoid biosynthetic pathway of *L. polystachyus* and identified the key genes in this pathway, *LpPAL3*, *LpCHS1*, *LpCHS2*, *LpCHI2*, and *LpF3H*. Besides, we deduced that the DTFs were involved in regulating the biosynthesis of different flavonoid metabolites and their regulatory patterns. The discovery preliminarily revealed the pathways and key genes of flavonoid biosynthesis in *L. polystachyus*, which provided a reference for further study on flavonoid biosynthesis.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c01125.

Primers used in this study, DAF mapped to the Ko00941 pathway in *L. polystachyus*, flavonoid-related DEGs in *L. polystachyus*, and DTFs and DEGs of *L. polystachyus* mapping to the homologous protein data of *Arabidopsis* in the STRING database (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Yuehong Long − *College of Life Sciences, North China University of Science and Technology, Tangshan 063210, China*; Email: longyh@ncst.edu.cn

Zhaobin Xing − *College of Life Sciences, North China University of Science and Technology, Tangshan 063210, China*; orcid.org/0000-0001-6810-4082; Email: xingzb@ncst.edu.cn

Authors

Duoduo Zhang − *College of Life Sciences, North China University of Science and Technology, Tangshan 063210, China*

Shuqing Wang − *Hospital of North China University of Science and Technology, Tangshan 063210, China*

Limei Lin − *College of Life Sciences, North China University of Science and Technology, Tangshan 063210, China*

Jie Zhang − *College of Life Sciences, North China University of Science and Technology, Tangshan 063210, China*

Minghui Cui − *College of Life Sciences, North China University of Science and Technology, Tangshan 063210, China*

Shuo Wang − *College of Life Sciences, North China University of Science and Technology, Tangshan 063210, China*

Xuelei Zhao − *College of Life Sciences, North China University of Science and Technology, Tangshan 063210, China*

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453

Jing Dong − *College of Life Sciences, North China University of Science and Technology, Tangshan 063210, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c01125

Funding

This work was supported by the Natural Science Foundation of Hebei Province (H2020209033) and Hebei Education Department-sponsored scientific research projects (ZD2019075).

Notes

The authors declare no competing financial interest.

Data Availability: The raw sequence data from this study have been deposited at the NCBI Sequence Read Archive (SRA), accession no. PRJNA787710. The mechanism raw sequence data from this study have been deposited at the MetaboLights, accession no. MTBLS4771.

REFERENCES

(1) Zhao, Y.; Li, X.; Zeng, X.; Huang, S.; Hou, S.; Lai, X. Characterization of phenolic constituents in *Lithocarpus polystachyus*. *Anal. Methods* **2014**, *6*, 1359−1363.

(2) Shang, A.; Liu, H.-Y.; Luo, M.; Xia, Y.; Yang, X.; Li, H.-Y.; Wu, D.-T.; Sun, Q.; Geng, F.; Gan, R.-Y. Sweet tea (*Lithocarpus polystachyus* rehd.) as a new natural source of bioactive dihydrochalcones with multiple health benefits. *Crit. Rev. Food Sci. Nutr.* **2022**, *62*, 917−934.

(3) Shang, A.; Luo, M.; Gan, R.-Y.; Xu, X.-Y.; Xia, Y.; Guo, H.; Liu, Y.; Li, H.-B. Effects of microwave-assisted extraction conditions on antioxidant capacity of sweet tea (*Lithocarpus polystachyus* Rehd.). *Antioxidants* **2020**, *9*, 678.

(4) Fang, H.-L.; Liu, M.-L.; Li, S.-Y.; Song, W.-Q.; Ouyang, H.; Xiao, Z.-P.; Zhu, H.-L. Identification, potency evaluation, and mechanism clarification of α-glucosidase inhibitors from tender leaves of *Lithocarpus polystachyus* Rehd. *Food Chem.* **2022**, *371*, 131128.

(5) Sun, Z.-G.; Li, Z.-N.; Zhang, J.-M.; Hou, X.-Y.; Yeh, S. M.; Ming, X. Recent Developments of Flavonoids with Various Activities. *Curr. Top. Med. Chem.* **2022**, *22*, 305−329.

(6) Liskova, A.; Samec, M.; Koklesova, L.; Samuel, S. M.; Zhai, K.; Al-Ishaq, R. K.; Abotaleb, M.; Nosal, V.; Kajo, K.; Ashrafizadeh, M.; Zarrabi, A.; Brockmueller, A.; Shakibaei, M.; Sabaka, P.; Mozos, I.; Ullrich, D.; Prosecky, R.; La Rocca, G.; Caprnda, M.; Büsselberg, D.; Rodrigo, L.; Kruzliak, P.; Kubatka, P. Flavonoids against the SARS-CoV-2 induced inflammatory storm. *Biomed. Pharmacother.* **2021**, *138*, 111430.

(7) Panche, A. N.; Diwan, A. D.; Chandra, S. R. Flavonoids: an overview. *J. Nutr. Sci.* **2016**, *5*, No. e47.

(8) Guo, Y.; Gao, C.; Wang, M.; Fu, F.-f.; El-Kassaby, Y. A.; Wang, T.; Wang, G. Metabolome and transcriptome analyses reveal flavonoids biosynthesis differences in Ginkgo biloba associated with environmental conditions. *Ind. Crops Prod.* **2020**, *158*, 112963.

(9) Zhao, M. R.; Li, J.; Zhu, L.; Chang, P.; Li, L. L.; Zhang, L. Y. Identification and characterization of MYB-bHLH-WD40 regulatory complex members controlling anthocyanidin biosynthesis in blueberry fruits development. *Genes* **2019**, *10*, 496.

(10) Deng, Y.; Lu, S. Biosynthesis and regulation of phenylpropanoids in plants. *Crit. Rev. Plant Sci.* **2017**, *36*, 257−290.

(11) Zhang, H.; Tao, H.; Yang, H.; Zhang, L.; Feng, G.; An, Y.; Wang, L. MdSCL8 as a negative regulator participates in ALA-induced *FLS1* to promote flavonol accumulation in apples. *Int. J. Mol. Sci.* **2022**, *23*, 2033.

(12) Aharoni, A.; De Vos, C. H. R.; Wein, M.; Sun, Z.; Greco, R.; Kroon, A.; Mol, J. N. M.; O'Connell, A. P. The strawberry *FaMYB1* transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. *Plant J.* **2001**, *28*, 319−332.

(13) Xie, X.-B.; Li, S.; Zhang, R.-F.; Zhao, J.; Chen, Y.-C.; Zhao, Q.; Yao, Y.-X.; You, C.-X.; Zhang, X.-S.; Hao, Y.-J. The bHLH transcription factor *MdbHLH3* promotes anthocyanin accumulation and fruit colouration in response to low temperature in apples. *Plant, Cell Environ.* **2012**, *35*, 1884−1897.

(14) Yang, J.; Huang, Y. Y.; Yang, Z.; Zhou, C.; Hu, X. J. Identification and quantitative evaluation of major sweet ingredients in sweet tea (*Lithocarpus polystachyus* Rehd.) based upon location, harvesting time, leaf age. *J. Chem. Soc. Pak.* **2018**, *40*, 158−164.

(15) Wei, M.; Tuo, Y.; Zhang, Y.; Deng, Q.; Shi, C.; Chen, X.; Zhang, X. Evaluation of two parts of *Lithocarpus polystachyus* Rehd. from different Chinese areas by multicomponent content determination and pattern recognition. *J. Anal. Methods Chem.* **2020**, *2020*, 1−10.

(16) Zhang, Y.; Lin, L.; Long, Y.; Guo, H.; Wang, Z.; Cui, M.; Huang, J.; Xing, Z. Comprehensive transcriptome analysis revealed the effects of the light quality, light intensity, and photoperiod on phlorizin accumulation in *Lithocarpus polystachyus* Rehd. *Forests* **2019**, *10*, 995.

(17) Li, S. Transcriptional control of flavonoid biosynthesis: fine-tuning of the MYB-bHLH-WD40 (MBW) complex. *Plant Signaling Behav.* **2014**, *9*, No. e27522.

(18) Xu, W.; Dubos, C.; Lepiniec, L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* **2015**, *20*, 176−185.

(19) Cao, Y.; Jia, H.; Xing, M.; Jin, R.; Grierson, D.; Gao, Z.; Sun, C.; Chen, K.; Xu, C.; Li, X. Genome-wide analysis of *MYB* gene family in Chinese bayberry (*Morella rubra*) and identification of members regulating flavonoid biosynthesis. *Front. Plant Sci.* **2021**, *12*, 691384.

(20) Salehi, B.; Fokou, P.; Sharifi-Rad, M.; Zucca, P.; Pezzani, R.; Martins, N.; Sharifi-Rad, J. The therapeutic potential of naringenin: a review of clinical trials. *Life Sci.* **2019**, *12*, 11.

(21) Yin, Y.-C.; Hou, J.-M.; Tian, S.-K.; Yang, L.; Zhang, Z.-X.; Li, W.-D.; Liu, Y. Overexpressing chalcone synthase (*CHS*) gene enhanced flavonoids accumulation in *Glycyrrhiza uralensis* hairy roots. *Bot. Lett.* **2019**, *167*, 219−231.

(22) Sun, W.; Shen, H.; Xu, H.; Tang, X.; Tang, M.; Ju, Z.; Yi, Y. Chalcone isomerase a key enzyme for anthocyanin biosynthesis in *Ophiorrhiza japonica*. *Front. Plant Sci.* **2019**, *10*, 865.

(23) Jiang, F.; Wang, J.-Y.; Jia, H.-F.; Jia, W.-S.; Wang, H.-Q.; Xiao, M. RNAi-mediated silencing of the flavanone 3-hydroxylase gene and its effect on flavonoid biosynthesis in strawberry fruit. *J. Plant Growth Regul.* **2013**, *32*, 182−190.

(24) Wei, Z.; Cheng, Y.; Zhou, C.; Li, D.; Gao, X.; Zhang, S.; Chen, M. Genome-wide identification of direct targets of the TTG1-bHLH-MYB complex in regulating trichome formation and flavonoid accumulation in *Arabidopsis Thaliana*. *Int. J. Mol. Sci.* **2019**, *20*, 5014.

(25) Lu, Y.; Bu, Y.; Hao, S.; Wang, Y.; Zhang, J.; Tian, J.; Yao, Y. MYBs affect the variation in the ratio of anthocyanin and flavanol in fruit peel and flesh in response to shade. *J. Photochem. Photobiol., B* **2017**, *168*, 40−49.

(26) Shan, X.; Li, Y.; Yang, S.; Yang, Z.; Qiu, M.; Gao, R.; Han, T.; Meng, X.; Xu, Z.; Wang, L.; Gao, X. The spatio-temporal biosynthesis of floral flavonols is controlled by differential phylogenetic MYB regulators in *Freesia hybrida*. *New Phytol.* **2020**, *228*, 1864−1879.

(27) Hichri, I.; Heppel, S. C.; Pillet, J.; Léon, C.; Czemmel, S.; Delrot, S.; Lauvergeat, V.; Bogs, J. The basic helix-loop-helix transcription factor MYC1 is involved in the regulation of the flavonoid biosynthesis pathway in grapevine. *Mol. Plant* **2010**, *3*, 509−523.

(28) Montefiori, M.; Brendolise, C.; Dare, A. P.; Lin-Wang, K.; Davies, K. M.; Hellens, R. P.; Allan, A. C. In the Solanaceae, a hierarchy of bHLHs confer distinct target specificity to the anthocyanin regulatory complex. *J. Exp. Bot.* **2015**, *66*, 1427−1436.

(29) Ma, D.; Constabel, C. P. MYB repressors as regulators of phenylpropanoid metabolism in plants. *Trends Plant Sci.* **2019**, *24*, 275−289.

(30) Wang, X. C.; Wu, J.; Guan, M. L.; Zhao, C. H.; Geng, P.; Zhao, Q. *Arabidopsis* MYB4 plays dual roles in flavonoid biosynthesis. *Plant J.* **2020**, *101*, 637−652.

(31) Czemmel, S.; Stracke, R.; Weisshaar, B.; Cordon, N.; Harris, N. N.; Walker, A. R.; Robinson, S. P.; Bogs, J. The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries. *Plant Physiol.* **2009**, *151*, 1513−1530.

https://doi.org/10.1021/acsomega.2c01125
ACS Omega 2022, 7, 19437−19453